# A Two-Step Composite Likelihood Approach with Data Fusion for Max-Stable Processes in Spatial Extremes Modelling

Hongwei Shang[1] Jun Yan[1,*]
Xuebin Zhang[2]

[1]Department of Statistics, University of Connecticut,
215 Glenbrook Rd. Unit 4120, Storrs, CT 06269, USA
[2]Environment Canada, Climate Data and Analysis,
4905 Dufferin Street, Downsview, Ontario M5H 5T4, Canada
jun.yan@uconn.edu



**Abstract**

Max-stable processes have recently been used in modeling spatial extremes. Due to unavailable full likelihood, inferences have to rely on composite likelihood constructed from bivariate or trivariate marginal densities. Since max-stable processes specify the dependence of block maxima, the data used for inference are block maxima from all sites. The point process approach or the peaks over threshold approach for univariate extreme value analysis, which uses more data and is preferred by practitioners, does not adapt easily to the spatial setting. To use more data in spatial extremes modeling, we propose a composite likelihood approach that fuses block maxima data with daily records. The procedure separates the estimation of marginal parameters and dependence parameters in two steps. The first step estimates marginal parameters using an independent composite likelihood from the point process approach with daily records. Given the marginal parameter estimates, the second step estimates the dependence parameters by a pairwise likelihood with block maxima. Based on the theory of inference functions, the asymptotic consistency and asymptotic normality can be established, and the limiting covariance matrix is estimated with a sandwich variance estimator. In a simulation study, the two-step estimator was found to be more efficient, and in some scenarios much more, than the existing composite likelihood approach based on block maxima data only. The practical utility is illustrated through the precipitation data at 20 sites in California over 55 years, and the two-step estimator gives much tighter confidence regions for risk measures of jointly defined events.

*Key words and phrases:* estimating functions, extreme value analysis, spatial dependence.

# 1 Introduction

Weather and climate extremes such as extreme temperature and precipitation have been associated with human health in both non-infectious and infectious diseases (Patz et al., 2003, 2005). Extremes data are often spatial in nature as data are recorded at a network of monitoring stations over time. Large scale atmospheric circulation such as El Niño may increase the occurrence of extreme weather events over large regions (e.g., Zhang et al., 2010). Rare events that occur at multiple locations within a very short time interval can cause more damage, consume more resources, and demand stronger emergency management. To mitigate losses and be better prepared, a holistic understanding of the dependence of extreme events in a spatial context is necessary. Although univariate extreme value modeling has been well developed (e.g., Coles, 2001), spatial extremes modeling has gained sharpened focus only recently (e.g., de Haan and Pereira, 2006; Buishand et al., 2008; Padoan et al., 2010; Davison and Gholamrezaee, 2012). Two recent reviews are Davison et al. (2012) and Bacro and Gaetan (2012), with the latter one focusing on spatial max-stable processes.

Max-stable processes are an important tool for spatial extremes modeling. A max-stable process is an infinite-dimensional extension of multivaraite extreme value distribution such that its every finite dimensional marginal distribution satisfies the max-stable property (de Haan, 1984). Specific parametric models have been applied in practice. Two most widely used models are the Smith model, also known as the Gaussian extreme value process (Smith, 1990), and the Schlather model, also known as the extremal Gaussian process (Schlather, 2002). Because the joint density is unavailable when four or more sites are involved, recent applications (Padoan et al., 2010; Davison and Gholamrezaee, 2012) relied on the composite likelihood approach (Lindsay, 1988) based on annual maxima data for inferences. The efficiency of the composite likelihood estimator can be improved if triplewise likelihood is used in place of pairwise likelihood, as shown by Genton et al. (2011) with the Smith model. Nevertheless, the trivariate marginal distribution has not been available for other max-stable process models.

For the univariate case, extreme value analysis can take advantage of more information than fitting the generalized extreme value (GEV) distribution to block maxima data (e.g., Coles, 2001). When daily records are available, two well known approaches are the peaks over threshold (POT) approach (Balkema and de Haan, 1974; Pickands, 1975) and, more generally, the point process approach (Pickands, 1971; Leadbetter et al., 1983). They are favored by practitioners in hydrology and climate research since they use more information in the data and perform better than the block maxima approach (e.g., Katz et al., 2002; Tanaka and Takara, 2002). A natural question is, can these approaches be applied in spatial extremes modeling to take fuller advantage of daily record? Unfortunately, a max-stable process model only specifies the dependence structure of the block maxima and says nothing about other exceedances over the thresholds. In fact, as pointed out by Falk and Michel (2009), application of the POT approach in multivariate extreme value modeling poses two main problems: 1) which distributions describe the exceedances, and 2) how exceedances are defined in a multivaraite setting. Investigation of these problems are still lively continuing (e.g., Rootzén and Tajvidi, 2006; Falk and Guillou, 2008; Falk et al., 2010).

Without resorting to a spatial version of the POT approach, can we make better use of the data if daily observations are available and improve the efficiency of parameter estimators? We propose a two-step composite likelihood approach for inferences in max-stable processes modeling with spatial extremes that fuses block maxima data with daily records from each site. The first step estimates marginal parameters using an independent composite likelihood for point processes with daily records data. Given the marginal parameter estimates, the second step estimates dependence parameters by a pairwise composite likelihood with block maxima data. The two-step approach has been studied recently for multivariate models to overcome the computational difficulty in maximum likelihood estimation (Zhao and Joe, 2005; Joe, 2005). A difference between our two-step approach and theirs is that we use different data in the two steps; the first step uses daily records while the second step uses annual maxima. With the separate estimation of marginal parameters and dependence parameters in two steps, the approach is easily grasped by practitioners. As expected, in our simulation study, the two-step estimator shows substantial gain in efficiency compared to the existing composite likelihood approach that uses only the block maxima data. Such gains can lead to much narrower confidence intervals for risk measures such as joint return levels.

The rest of the article is organized as follows. The spatial max-stable process model defined by all univariate marginal distributions and a spatial dependence structure is introduced in Section 2. In Section 3, we present details of the two-step composite likelihood approach, the asymptotic properties of the estimator, and how to estimate the limiting variance. A simulation study is reported in Section 4 to assess the efficiency gain of the two-step approach in comparison to the current practice based on block maxima data only. The proposed method is applied to the precipitation data from 20 sites in California over 55 years in Section 5, providing more compact confidence regions for joint return levels. Section 6 concludes with some discussion.

# 2 Spatial Max-stable Process Model

Max-stable processes are extensions of the multivariate extreme value distribution to the infinite dimensional case. Without loss of generality, max-stable processes are often presented with unit Fréchet margins with distribution function $F(z) = \exp(-1/z)$, $z > 0$. Such max-stable processes are known as simple max-stable processes. A simple max-stable process $Z(s)$ on a spatial domain $D$ with unit Fréchet margins has all finite dimensional marginal distributions satisfying the max-stablility property:

$$\Pr\{Z(x_1) \le kz_1, \dots, Z(x_p) \le kz_p\}^k = \Pr\{Z(x_1) \le z_1, \dots, Z(x_p) \le z_p\},$$

for any $p$ sites $\{x_1, \dots, x_p\} \subset D$, all $z_1 > 0, \dots, z_p > 0$ and $k \in \mathbb{N}$. The max-stablity property is equivalently specifying that all $p$-order marginal distributions are multivariate extreme value distributions.

Two classes of spectral representations of max stable processes have been proposed by, respectively, de Haan (1984) and Schlather (2002), both of which can be put into a unified

form (Davison et al., 2012). Let $\Pi$ be a Poisson process on $R_+$ with intensity $\mathrm{d}u/u^2$. Let $W_\xi(x)$, $x \in D$, $\xi > 0$, be independent copies of a non-negative stationary process $W(x)$ with $E\{W(x)\} = 1$ for all $x \in D$. Then,

$$Z(x) = \max_{\xi \in \Pi} \xi W_\xi(x), \qquad x \in D,$$

is a stationary simple max-stable process.

Parametric models for max-stable processes can be constructed by choosing different random process $W(x)$. For example, a class of rainfall storm model is obtained by letting $W(x) = g(x-\xi)$, where $\xi$ is a random point in $D$, and $g$ is a density function (de Haan, 1984; Smith, 1990). This model has a nice interpretation that $W(x)$ is the rainfall at location $x$ from a storm with center $\xi$ and shape $g$. If $g$ is a normal density with mean zero and covariance matrix $\Sigma$, the resulting process $Z(s)$ is the Smith model, also known as the Gaussian extreme value process. This model receives a great interest due to its mathematical tractability and nice interpretation (Smith, 1990; Coles, 1993; de Haan and Pereira, 2006; Padoan et al., 2010; Genton et al., 2011). The bivariate marginal distribution function at two sites $x_1$ and $x_2$ can be shown to be

$$\Pr[Z(x_1) \leq z_1, Z(x_2) \leq z_2] = \exp\left\{ -\frac{1}{z_1}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_1}{z_2}\right)\right\},$$

where $\Phi$ is the standard normal cumulative distribution function, $a^2 = \Delta x^\top\, \Sigma^{-1} \Delta x$, and $\Delta x = x_1 - x_2$. The density function can be obtained by differentiating the distribution function.

A max-stable process model for spatial extremes is decomposed into two parts: marginal distributions and the spatial dependence structure. Let $M(s)$ be the extremal variable at site $s$ in a domain $D \subset R^2$. Let $X(s)$ be the covariate vector available at site $s$. The marginal models are GEV distributions:

$$M(s) \sim \mathrm{GEV}\big(\mu(s), \sigma(s), \xi(s)\big), \tag{1}$$

where $\mu(s)$, $\sigma(s)$, and $\xi(s)$ are the location, scale, and shape parameters of the GEV distribution at site $s$, respectively. The covariate vector is incorporated into the parameters through

$$\begin{aligned} g_\mu\big(\mu(s)\big) &= X^\top(s)\beta_\mu, \\ g_\sigma\big(\sigma(s)\big) &= X^\top(s)\beta_\sigma, \\ g_\xi\big(\xi(s)\big) &= X^\top(s)\beta_\xi, \end{aligned}$$

where $g_\mu$, $g_\sigma$, and $g_\xi$ are known link functions for $\mu$, $\sigma$, and $\xi$, respectively, and $\beta^\top = (\beta_\mu^\top, \beta_\sigma^\top, \beta_\xi^\top)$ is the vector containing all marginal parameters. The spatial dependence is characterized by a max-stable process (MSP) with dependence parameter $\theta$:

$$F^{-1}\big\{G_s\big(M(s); \beta\big)\big\} \sim \mathrm{MSP}(\theta), \tag{2}$$

where $F$ is the distribution function of unit Fréchet, and $G_s(\cdot;\beta)$ is the distribution function of $\text{GEV}\big(\mu(s),\sigma(s),\xi(s)\big)$ with parameter vector $\beta$. If the Smith model is used, for instance, $\theta$ would contain all three parameters in $\Sigma$, $(\sigma_{11},\sigma_{12},\sigma_{22})$. The whole model is characterized by $\eta^\top = (\beta^\top, \theta^\top)$, which concatenates the marginal parameters and the dependence parameters.

Since the full likelihood of $Z(s)$ is unavailable, the current practice in parameter estimation is the composite likelihood approach constructed from bivariate marginal densities (e.g., Padoan et al., 2010). The daily records from each site $s$ do not contribute to the estimation except the maximum from block (year), which may otherwise increase the efficiency in marginal parameter estimation and in turn improve the efficiency in dependence parameter estimation as well. This motivates our two-step approach.

# 3 Two-Step Approach

Suppose that we observe the full record of each block with block size $m$ at $S$ sites over $n$ blocks (e.g. years, months). Let $\mathbf{Y_{s,t}}$ be the whole observation vector at site $s$ in block $t$, $s=1,\dots,S$, $t=1,\dots,n$. Let $Y_{s,t,k}$ be the $k$th observation within block $t$ at site $s$, $k=1,\dots,m$ and $M_{s,t}$ be the block maximum. The block maxima data $\mathbf{M}=\{M_{s,t}: s=1,\dots,S; t=1,\dots,n\}$ are observations from the model defined in (1) and (2). Let $X(s)$ be the covariate vector at site $s$; examples are longitude, latitude, and elevation. For simplicity, we assume that there is no temporal dependence; temporal dependence can be handled with declustering in practice.

Our two-step composite likelihood procedure estimates marginal parameters in the first step and the dependence parameters in the second step.

**Step 1** The first step is based on an independent likelihood constructed from the point process approach for univariate extreme value analysis, utilizing all full block data but ignoring the spatial dependence. Let $\mathbf{Y}=\{\mathbf{Y}_{s,t}: s=1,\dots,S; t=1,\dots,n\}$. Let $u_s$ be the threshold chosen for site $s$, $s=1,\dots,S$. The independent loglikelihood has the form

$$l_1(\beta;\mathbf{Y}) = \sum_{t=1}^{n}\sum_{s=1}^{S} \ell_{1t,s}(\beta;\mathbf{Y}_{s,t}), \tag{3}$$

where

$$\begin{aligned}\ell_{1t,s}(\beta;\mathbf{Y}_{s,t}) = &- \left[1+\xi_s\left(\frac{u_s-\mu_s}{\sigma_s}\right)\right]^{-1/\xi_s} \\ &+ \sum_{k:Y_{s,t,k}>u_s}\left[-\log\sigma_s - \left(\frac{1}{\xi_s}+1\right)\log\left\{1+\xi_s\left(\frac{Y_{s,t,k}-\mu_s}{\sigma_s}\right)\right\}\right].\end{aligned}$$

The contribution to the independent loglikelihood from site $s$, $\sum_{t=1}^{n}\ell_{1t,s}$, is simply loglikelihood of the point process approach in a univariate extreme value analysis (Smith, 1989). Since we assume indepdence from block to block, the contribution from block $t$ is $\ell_{1t}=\sum_{s=1}^{S}\ell_{1t,s}$. The maximizer of (3), $\hat\beta_n$, is the estimator of $\beta$.

**Step 2** The second step uses only block maxima and estimates dependence parameters based on a pairwise composite likelihood, with marginal parameters replaced by the estimates from the first step. The dependence parameters are estimated by maximizing a pairwise likelihood based on block maxima with $\beta$ fixed at $\hat{\beta}_n$. Let $f_{ij}(\cdot;\theta,\beta)$ be the bivariate marginal density of the MSP in (2) at site $i$ and $j$, with dependence parameters $\theta$ and marginal parameters $\beta$. Define a log pairwise composite likelihood

$$l_2(\theta;\hat{\beta}_n,\mathbf{M}) = \sum_{t=1}^{n} \ell_{2t}(\theta;\hat{\beta}_n, M_{s,t} : s = 1,\dots,S), \tag{4}$$

where the contribution from block $t$ is

$$\ell_{2t}(\theta;\beta, M_{s,t} : s = 1,\dots,S) = \sum_{i=1}^{S-1}\sum_{j=i+1}^{S} \log f_{i,j}\big((M_{i,t}, M_{j,t});\theta,\beta\big).$$

Our estimator for $\theta$, $\hat{\theta}_n$, is the maximizer of (4).

The asymptotic properties of the two-step estimator $\hat{\eta}_n^\top = (\hat{\beta}_n^\top, \hat{\theta}_n^\top)$ can be derived with general theory of estimating functions (Godambe, 1991). Let $\psi_{1t}(\beta) = \partial\ell_{1t}/\partial\beta$. Let $\psi_{2t}(\beta,\theta) = \partial\ell_{2t}/\partial\theta$. The estimator $\hat{\eta}_n$ is the solution to the estimating equations

$$\sum_{t=1}^{n} \psi_t(\eta) = 0, \tag{5}$$

where

$$\psi_t(\eta) = \begin{pmatrix} \psi_{1t}(\beta) \\ \psi_{2t}(\beta,\theta) \end{pmatrix}.$$

Under mild regularity conditions, as $n \to \infty$, the solution $\hat{\eta}_n$ to (5) is consistent to the true parameter vector $\eta_0$, and

$$\sqrt{n}(\hat{\eta}_n - \eta_0) \to N(0,\Omega),$$

where $\Omega = A^{-1}B(A^{-1})^\top$ is the inverse of the Godambe information matrix, with $A = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} \partial\psi_t(\eta)/\partial\eta^\top$, and $B = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} \psi_t(\eta)\psi_t^\top(\eta)$.

With independent replicates at the block level, $\Sigma$ can be easily estimated with the sample versions of $A$ and $B$. We estimate $A$ with

$$A_n = \frac{1}{n}\sum_{t=1}^{n} \begin{pmatrix} \partial\psi_{1t}(\hat{\beta}_n)/\partial\beta^\top & \\ \partial\psi_{2t}(\hat{\beta}_n,\hat{\theta}_n)/\partial\beta^\top & \partial\psi_{2t}(\hat{\beta}_n,\hat{\theta}_n)/\partial\theta^\top \end{pmatrix},$$

where the triangular form comes from the fact that the first step does not involve the dependence parameters. This matrix involves the second-order derivatives of the log composite likelihood, which can be difficult to obtain in practice. An alternative formula that invloves only the first-order derivatives of the log composite likelihoods is given in the Appendix. For $B$, we estimate it with

$$B_n = \frac{1}{n}\sum_{t=1}^{n} \begin{pmatrix} \psi_{1t}(\hat{\beta}_n)\psi_{1t}^\top(\hat{\beta}_n) & \psi_{1t}(\hat{\beta}_n)\psi_{2t}^\top(\hat{\beta}_n,\hat{\theta}_n) \\ \psi_{2t}(\hat{\beta}_n,\hat{\theta}_n)\psi_{1t}^\top(\hat{\beta}_n) & \psi_{2t}(\hat{\beta}_n,\hat{\theta}_n)\psi_{2t}^\top(\hat{\beta}_n,\hat{\theta}_n) \end{pmatrix}.$$

Our estimator of $\Omega$ is then $\hat{\Omega}_n = A_n^{-1} B_n (A_n^{-1})^\top$. The estimation of $B$ in this problem is straightforward because we have replicates, unlike in an usual spatial data setting where there is only one replicate and some bootstrap procedure would need to be developed (Heagerty and Lele, 1998; Heagerty and Lumley, 2000).

# 4 Simulation Study

A simulation study was conducted to investigate the performance of the two-step approach using daily record in comparison with the existing pairwise composite likelihood approach using block maxima only. It also assessed the validity of the sandwich variance estimator of the asymptotic variance of the estimators from the two-step approach. The study region is confined to be $[-5, 5]^2$. Factors of our experiment are: the marginal model, spatial dependence structure, the number of sites $S$, and the sample size $n$ (usually corresponding to the number of years in a real dataset).

The marginal distribution of block maxima at each site $s$ is a GEV distribution with location $\mu_s$, scale $\sigma_s$, and shape $\xi_s$. Let $X_1(s)$ and $X_2(s)$ denote the latitude and longitude of site $s$. We considered first a model of the form

$$\begin{cases} \mu_s = \beta_{\mu,0} + \beta_{\mu,1} X_1(s) + \beta_{\mu,2} X_2(s), \\ \sigma_s = \beta_{\sigma,0}, \\ \xi_s = \beta_{\xi,0}, \end{cases}$$

where $\beta_{\mu,0} = 5$, $\beta_{\mu,1} = -0.5$, $\beta_{\mu,2} = 1$, $\beta_{\sigma,0} = 2.5$, and $\beta_{\xi,0} = 0.2$. The Smith model was used for the spatial dependence structure. Two dependence levels were considered: weak anisotropic dependence with $\Sigma = \Sigma_1 = 4Q$ and moderate anisotropic dependence with $\Sigma = \Sigma_2 = 16Q$, where $Q$ is a bivariate correlation matrix with correlation coefficient 0.5. To make the simulation realistic, we considered the number of sites $S \in \{25, 49\}$ and the sample size $n \in \{20, 50, 100\}$. For each $S$, the sites form a square grid over $[-5, 5]^2$ (Figure 1).

For each scenario, 1000 datasets of daily observations were generated. Note that the daily observations need to be generated with care such that the annual maxima follow the max-stable process models specified by the marginal GEV models and the Smith dependence structure. For each day, we generated a realization from the simple Smith model with specified dependence level and divided the realization by 365. Assuming independence from day to day (in a real data analysis declustering would be done to account for temporal dependence), the componentwise maximum vector at all sites over 365 days follow the simple Smith model by the max-stability property. The daily series at each site were transformed to have the specified GEV marginal distributions for the annual maxima.

For a given dataset, the R package `SpatialExtremes` (Ribatet, 2011) was used to obtain the pairwise likelihood estimator based on annual maxima data. These estimates were used as the starting values in the two-step approach, where the general purpose optimizer `optim` in R was used to maximize the composite likelihoods in the two steps. The threshold in the first step of two-step approach was chosen to be the 95th sample percentile at each site.

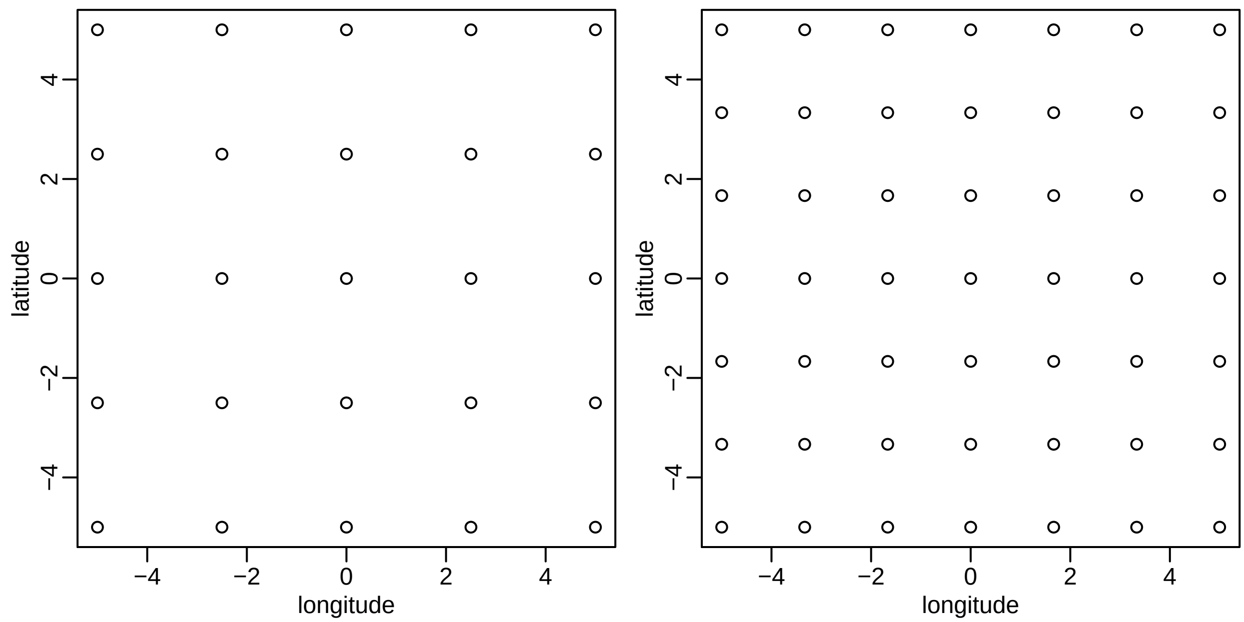


Figure 1: Sites generated for simulation study. Left panel: $S = 25$; right panel: $S = 49$.

The empirical mean squared error (MSE) of the estimators from the two-step approach and pairwise likelihood approach based on the 1000 replicates are summarized in Table 1. As expected, the MSEs from both approaches decrease as sample size $n$ increases for both marginal and dependence parameters. As the dependence level gets stronger from $\Sigma_1$ to $\Sigma_2$, the MSEs from both approaches increase, and the magnitude of the change is more for dependence parameters than for marginal parameters. The number of sites has little affect on the MSEs for both approaches. For a given combination of $(n, S, \Sigma)$, the two-step approach always gives smaller MSE than the pairwise likelihood approach. The difference is more substantial for marginal parameters than for dependence parameters, especially with the regression coefficients in the location parameter, in which cases, the MSE of the two-step estimator is about 2% of that of the pairwise likelihood estimator.

For a better inspection, Table 2 reports the relative efficiency of the two approaches for each parameter — the ratio of the empirical MSE from the two-step approach ($M_1$) to that from the block maxima pairwise likelihood approach ($M_2$). Overall, the efficiency gains of the two-step approach are large, with the relative efficiency varying from 1.7% to 87%. The two-step approach improves substantially for the marginal parameters. In particular, we observe a striking efficiency gain (with relative efficiency between 1.7% and 3.9%) in $\beta_{\mu,1}$ and $\beta_{\mu,2}$, the coefficients of latitude and longitude in the location parameter. It is also of interest to look at $\beta_{\xi,0}$, since the shape parameter $\xi$ governs the tail behaviour of the GEV distribution and plays an important role in predicting return levels. The relative efficiency for $\beta_{\xi,0}$ is between 16% and 24% for all scenarios, which also indicates substantial gain. For the dependence parameters, the efficiency gain is appreciable with relative efficiency ranging

Table 1: Empirical MSE for model 1 based on two approaches ($M_1$: the existing pairwise likelihood approach using block maxima data; $M_2$: the two-step approach using data fusion).

| | | $n=20$ | | | | $n=50$ | | | | $n=100$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\Sigma_1$ | | $\Sigma_2$ | | $\Sigma_1$ | | $\Sigma_2$ | | $\Sigma_1$ | | $\Sigma_2$ | |
| $S$ | | 25 | 49 | 25 | 49 | 25 | 49 | 25 | 49 | 25 | 49 | 25 | 49 |
| $\sigma_{11}$ | $M_1$ | 1.30 | 1.09 | 38.94 | 44.02 | 0.52 | 0.45 | 16.56 | 16.92 | 0.24 | 0.22 | 7.88 | 8.39 |
| | $M_2$ | 1.00 | 0.82 | 29.60 | 34.54 | 0.39 | 0.32 | 12.42 | 12.91 | 0.18 | 0.16 | 5.89 | 6.01 |
| $\sigma_{12}$ | $M_1$ | 0.70 | 0.60 | 22.48 | 21.92 | 0.28 | 0.22 | 8.05 | 8.60 | 0.12 | 0.11 | 3.87 | 3.86 |
| | $M_2$ | 0.60 | 0.51 | 19.04 | 18.67 | 0.24 | 0.18 | 6.87 | 7.33 | 0.11 | 0.10 | 3.27 | 3.08 |
| $\sigma_{22}$ | $M_1$ | 1.34 | 1.18 | 50.07 | 46.51 | 0.54 | 0.42 | 16.85 | 16.52 | 0.22 | 0.22 | 7.84 | 7.98 |
| | $M_2$ | 0.97 | 0.86 | 40.06 | 35.96 | 0.40 | 0.30 | 12.94 | 13.16 | 0.17 | 0.16 | 5.88 | 5.65 |
| $\beta_{\mu,0}(\times 10^{-2})$ | $M_1$ | 7.49 | 7.85 | 16.27 | 16.73 | 2.68 | 3.23 | 7.47 | 7.38 | 1.43 | 1.42 | 3.28 | 3.37 |
| | $M_2$ | 5.72 | 5.85 | 11.27 | 12.55 | 2.24 | 2.34 | 5.13 | 4.99 | 1.05 | 1.12 | 2.28 | 2.28 |
| $\beta_{\mu,1}(\times 10^{-4})$ | $M_1$ | 18.11 | 19.49 | 26.94 | 29.68 | 7.10 | 7.61 | 11.78 | 12.00 | 3.41 | 3.92 | 4.81 | 5.31 |
| | $M_2$ | 0.40 | 0.43 | 0.91 | 0.54 | 0.16 | 0.18 | 0.22 | 0.22 | 0.08 | 0.09 | 0.10 | 0.12 |
| $\beta_{\mu,2}(\times 10^{-4})$ | $M_1$ | 17.74 | 20.31 | 28.56 | 31.11 | 7.00 | 7.11 | 10.46 | 11.53 | 3.55 | 3.73 | 4.71 | 5.58 |
| | $M_2$ | 0.39 | 0.42 | 1.11 | 0.53 | 0.16 | 0.16 | 0.20 | 0.23 | 0.08 | 0.08 | 0.10 | 0.10 |
| $\beta_{\sigma,0}(\times 10^{-2})$ | $M_1$ | 4.02 | 4.29 | 9.70 | 9.95 | 1.65 | 1.74 | 4.04 | 3.91 | 0.79 | 0.80 | 1.86 | 2.08 |
| | $M_2$ | 3.20 | 3.36 | 6.84 | 7.22 | 1.33 | 1.29 | 3.05 | 2.83 | 0.59 | 0.63 | 1.38 | 1.40 |
| $\beta_{\xi,0}(\times 10^{-3})$ | $M_1$ | 4.84 | 4.64 | 8.89 | 8.58 | 2.03 | 1.88 | 3.15 | 3.15 | 0.96 | 0.92 | 1.64 | 1.72 |
| | $M_2$ | 0.78 | 0.83 | 1.84 | 1.81 | 0.35 | 0.33 | 0.72 | 0.68 | 0.18 | 0.19 | 0.39 | 0.39 |

Table 2: Relative efficiency in MSE for the pairwise likelihood approach using block maxima data relative to the two-step approach using data fusion under marginal model 1.

| $n$ | $\Sigma$ | $S$ | $\sigma_{11}$ | $\sigma_{12}$ | $\sigma_{22}$ | $\beta_{\mu,0}$ | $\beta_{\mu,1}$ | $\beta_{\mu,2}$ | $\beta_{\sigma,0}$ | $\beta_{\xi,0}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 20 | $\Sigma_1$ | 25 | 0.77 | 0.85 | 0.73 | 0.76 | 0.022 | 0.022 | 0.79 | 0.16 |
| | | 49 | 0.75 | 0.85 | 0.73 | 0.75 | 0.022 | 0.021 | 0.78 | 0.18 |
| | $\Sigma_2$ | 25 | 0.76 | 0.85 | 0.80 | 0.69 | 0.034 | 0.039 | 0.71 | 0.21 |
| | | 49 | 0.78 | 0.85 | 0.77 | 0.75 | 0.018 | 0.017 | 0.73 | 0.21 |
| 50 | $\Sigma_1$ | 25 | 0.75 | 0.84 | 0.73 | 0.84 | 0.022 | 0.024 | 0.80 | 0.17 |
| | | 49 | 0.72 | 0.82 | 0.71 | 0.72 | 0.023 | 0.023 | 0.74 | 0.17 |
| | $\Sigma_2$ | 25 | 0.75 | 0.85 | 0.77 | 0.69 | 0.018 | 0.019 | 0.75 | 0.23 |
| | | 49 | 0.76 | 0.85 | 0.80 | 0.68 | 0.018 | 0.020 | 0.72 | 0.22 |
| 100 | $\Sigma_1$ | 25 | 0.76 | 0.87 | 0.75 | 0.74 | 0.024 | 0.022 | 0.76 | 0.19 |
| | | 49 | 0.75 | 0.85 | 0.74 | 0.79 | 0.023 | 0.022 | 0.78 | 0.20 |
| | $\Sigma_2$ | 25 | 0.75 | 0.84 | 0.75 | 0.70 | 0.021 | 0.022 | 0.74 | 0.24 |
| | | 49 | 0.72 | 0.80 | 0.71 | 0.67 | 0.022 | 0.018 | 0.67 | 0.23 |

from 71% to 85%. The gain here can be explained by the fact that the marginal parameters are estimated more precisely with the daily data in the first step.

To examine the bias and the performance of the sandwich variance estimator of the two-step approach, we summarize in Table 3 the average of biases, the empirical standard error, and the average of the standard error from the sandwich variance estimator based on 1000 replicates. To save space, only results for $n \in \{20, 50\}$ are reported; the case of $n = 100$ is omitted because the results are already good at $n = 50$. The biases are very small compared to the true values for all parameters. The average standard errors are slightly smaller than the empirical standard errors for all parameter estimates with $n = 20$. Consequently, the empirical coverage rates of 95% confidence intervals are slightly smaller than the nominal rate for some parameters. Nevertheless, with sample size $n = 50$, the agreement between the empirical standard errors and the average standard errors improves, and the empirical coverage rates of 95% confidence intervals are reasonably close to 95%.

We also considered a second marginal model of the form

$$\begin{cases} \mu_s = \beta_{\mu,0} + \beta_{\mu,1} X_1(s) + \beta_{\mu,2} X_2(s) \\ \sigma_s = \beta_{\sigma,0} + \beta_{\sigma,1} X_1(s) + \beta_{\sigma,2} X_2(s) \\ \xi_s = \beta_{\xi,0}, \end{cases}$$

where $\beta_{\mu,0} = 5$, $\beta_{\mu,1} = -0.5$, $\beta_{\mu,2} = 1$, $\beta_{\sigma,0} = 2.5$, $\beta_{\sigma,1} = 0.2$, $\beta_{\sigma,2} = -0.2$, and $\beta_{\xi,0} = 0.2$. This model had scale parameter depends on the covariates, and generated data with much higher variation than the first model. The relative efficiencies in MSE are summarized in Table 4, ranging from 22% to 100%. The results indicate that the two-step approach is superior to the pairwise likelihood approach for all parameters in this model too, especially

Table 3: Bias, empirical standard errors (ESE), average standard errors (ASE) from sandwich estimators, and empirical coverage prabability (CP) of 95% confidence intervals for the model parameters (Par) based on 1000 replicates with marginal model 1.

| | | | $n = 20$ | | | | $n = 50$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\Sigma$ | $S$ | Par | bias | ESE | ASE | CP | bias | ESE | ASE | CP |
| $\Sigma_1$ | 25 | $\sigma_{11}$ | 0.08681 | 0.994 | 0.852 | 89.1 | 0.01726 | 0.626 | 0.562 | 91.1 |
| | | $\sigma_{12}$ | 0.07298 | 0.768 | 0.673 | 91.4 | 0.03012 | 0.488 | 0.441 | 92.5 |
| | | $\sigma_{22}$ | 0.08737 | 0.984 | 0.854 | 89.4 | 0.03365 | 0.630 | 0.566 | 91.0 |
| | | $\beta_{\mu,0}$ | 0.00128 | 0.239 | 0.231 | 93.5 | $-0.00407$ | 0.150 | 0.147 | 93.3 |
| | | $\beta_{\mu,1}$ | 0.00033 | 0.006 | 0.007 | 95.9 | 0.00006 | 0.004 | 0.004 | 96.1 |
| | | $\beta_{\mu,2}$ | 0.00016 | 0.006 | 0.007 | 95.1 | $-0.00005$ | 0.004 | 0.004 | 95.3 |
| | | $\beta_{\sigma,0}$ | $-0.00527$ | 0.179 | 0.171 | 92.5 | $-0.00432$ | 0.115 | 0.109 | 92.1 |
| | | $\beta_{\xi,0}$ | $-0.00821$ | 0.027 | 0.025 | 89.9 | $-0.00624$ | 0.018 | 0.016 | 91.3 |
| | 49 | $\sigma_{11}$ | 0.05451 | 0.904 | 0.803 | 89.2 | $-0.00909$ | 0.568 | 0.536 | 92.5 |
| | | $\sigma_{12}$ | 0.06876 | 0.712 | 0.629 | 89.3 | $-0.00275$ | 0.425 | 0.411 | 93.3 |
| | | $\sigma_{22}$ | 0.09252 | 0.922 | 0.819 | 90.2 | $-0.00873$ | 0.544 | 0.539 | 93.7 |
| | | $\beta_{\mu,0}$ | $-0.00901$ | 0.242 | 0.241 | 93.5 | 0.00409 | 0.153 | 0.153 | 94.8 |
| | | $\beta_{\mu,1}$ | 0.00001 | 0.007 | 0.007 | 95.0 | 0.00019 | 0.004 | 0.004 | 95.1 |
| | | $\beta_{\mu,2}$ | 0.00005 | 0.007 | 0.007 | 94.6 | 0.00003 | 0.004 | 0.004 | 96.5 |
| | | $\beta_{\sigma,0}$ | $-0.01379$ | 0.183 | 0.177 | 91.7 | $-0.00284$ | 0.114 | 0.114 | 93.7 |
| | | $\beta_{\xi,0}$ | $-0.00964$ | 0.027 | 0.026 | 89.9 | $-0.00667$ | 0.017 | 0.017 | 91.4 |
| $\Sigma_2$ | 25 | $\sigma_{11}$ | 0.81744 | 5.382 | 5.072 | 91.8 | 0.61117 | 3.473 | 3.353 | 94.5 |
| | | $\sigma_{12}$ | 0.63898 | 4.319 | 3.862 | 91.8 | 0.25649 | 2.609 | 2.478 | 94.0 |
| | | $\sigma_{22}$ | 1.31173 | 6.195 | 5.229 | 91.4 | 0.51189 | 3.562 | 3.338 | 93.2 |
| | | $\beta_{\mu,0}$ | $-0.01700$ | 0.335 | 0.347 | 94.7 | 0.01510 | 0.226 | 0.221 | 94.8 |
| | | $\beta_{\mu,1}$ | 0.00018 | 0.010 | 0.008 | 94.8 | 0.00014 | 0.005 | 0.005 | 94.8 |
| | | $\beta_{\mu,2}$ | 0.00029 | 0.011 | 0.007 | 95.6 | $-0.00034$ | 0.004 | 0.005 | 95.6 |
| | | $\beta_{\sigma,0}$ | $-0.02125$ | 0.261 | 0.259 | 91.9 | 0.01115 | 0.174 | 0.167 | 94.5 |
| | | $\beta_{\xi,0}$ | $-0.01214$ | 0.041 | 0.039 | 90.5 | $-0.00467$ | 0.026 | 0.025 | 93.5 |
| | 49 | $\sigma_{11}$ | 0.92824 | 5.806 | 5.116 | 91.5 | 0.50904 | 3.558 | 3.311 | 93.8 |
| | | $\sigma_{12}$ | 0.45541 | 4.299 | 3.755 | 90.8 | 0.29361 | 2.693 | 2.428 | 92.3 |
| | | $\sigma_{22}$ | 0.92646 | 5.927 | 5.105 | 90.1 | 0.54582 | 3.588 | 3.329 | 93.1 |
| | | $\beta_{\mu,0}$ | $-0.00363$ | 0.354 | 0.361 | 93.5 | 0.00360 | 0.224 | 0.226 | 95.3 |
| | | $\beta_{\mu,1}$ | 0.00002 | 0.007 | 0.008 | 95.7 | 0.00003 | 0.005 | 0.005 | 94.7 |
| | | $\beta_{\mu,2}$ | 0.00031 | 0.007 | 0.008 | 95.3 | 0.00000 | 0.005 | 0.005 | 94.8 |
| | | $\beta_{\sigma,0}$ | $-0.01033$ | 0.269 | 0.269 | 92.4 | 0.00066 | 0.168 | 0.170 | 94.7 |
| | | $\beta_{\xi,0}$ | $-0.01041$ | 0.041 | 0.040 | 91.2 | $-0.00627$ | 0.025 | 0.026 | 93.7 |

Table 4: Relative efficiency of MSE of the pairwise likelihood approach using block maxima data relative to the two-step approach using data fusion under marginal model 2.

| $n$ | $\Sigma$ | $S$ | $\sigma_{11}$ | $\sigma_{12}$ | $\sigma_{22}$ | $\beta_{\mu,0}$ | $\beta_{\mu,1}$ | $\beta_{\mu,2}$ | $\beta_{\sigma,0}$ | $\beta_{\sigma,1}$ | $\beta_{\sigma,2}$ | $\beta_{\xi,0}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | $\Sigma_1$ | 25 | 0.74 | 0.73 | 0.71 | 0.91 | 0.63 | 0.65 | 1.00 | 0.35 | 0.33 | 0.23 |
| | | 49 | 0.76 | 0.73 | 0.68 | 0.85 | 0.58 | 0.63 | 0.91 | 0.31 | 0.31 | 0.23 |
| | $\Sigma_2$ | 25 | 0.69 | 0.74 | 0.79 | 0.82 | 0.66 | 0.71 | 0.88 | 0.44 | 0.45 | 0.26 |
| | | 49 | 0.75 | 0.74 | 0.76 | 0.84 | 0.65 | 0.67 | 0.86 | 0.42 | 0.43 | 0.27 |
| 50 | $\Sigma_1$ | 25 | 0.75 | 0.82 | 0.77 | 0.96 | 0.62 | 0.63 | 0.95 | 0.36 | 0.32 | 0.22 |
| | | 49 | 0.73 | 0.73 | 0.68 | 0.90 | 0.63 | 0.67 | 0.92 | 0.34 | 0.35 | 0.23 |
| | $\Sigma_2$ | 25 | 0.75 | 0.82 | 0.79 | 0.87 | 0.69 | 0.74 | 0.94 | 0.49 | 0.50 | 0.29 |
| | | 49 | 0.75 | 0.76 | 0.73 | 0.77 | 0.60 | 0.60 | 0.84 | 0.43 | 0.41 | 0.26 |
| 100 | $\Sigma_1$ | 25 | 0.81 | 0.87 | 0.76 | 0.87 | 0.57 | 0.55 | 0.93 | 0.34 | 0.30 | 0.25 |
| | | 49 | 0.75 | 0.84 | 0.76 | 0.88 | 0.61 | 0.65 | 0.89 | 0.31 | 0.32 | 0.24 |
| | $\Sigma_2$ | 25 | 0.77 | 0.83 | 0.78 | 0.80 | 0.68 | 0.65 | 0.85 | 0.47 | 0.41 | 0.29 |
| | | 49 | 0.74 | 0.79 | 0.72 | 0.82 | 0.70 | 0.68 | 0.83 | 0.49 | 0.42 | 0.28 |

for the shape parameter (20–29%) and the regression coefficients of latitude and longitude in the scale (31–50%) parameter. In contrast to model 1, the efficiency gain for $\beta_{\mu,1}$ and $\beta_{\mu,2}$ of model 2 is not as large as that of model 1, This may be due to the simplicity of model 1, in which latitude and longitude only appear in $\mu_s$.

# 5 Extreme Winter Precipitation in California

Daily precipitation records at all monitoring stations in California were extracted from the second version of the Global Historical Climatology Network (GHCN), compiled and quality controlled at the National Climatic Data Center of the National Oceanic and Atmospheric Administration (available online at http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/). The same data source was used in Shang et al. (2011). The raw data contains daily rainfall observations at 230 sites in California starting from 1878. As most precipitations in California occur in winters, only the winter extreme precipitation is considered. The winter season here spans from December 1st to March 31st in the following year (Zhang et al., 2010). Due to missing data, the maximum observation in a given year at a given site was considered as the annual maximum only if there were no more than 5% missing daily observations during the winter season. In our analysis, we used a balanced data set consisting of daily winter precipitation from 1948 to 2002 for 20 sites (see Figure 2).

We propose a spatial max-stable process model for the California data. Three explanatory variables are available for the marginal parameters. Let $X^\top(s) = \{X_1(s), X_2(s), X_3(s)\}$ be the longitude, latitude, and elevation at site $s$, respectively. The first two are in degrees, and the elevation is in 100 meters. They are centered at the values at San Francisco ($-122.38$,

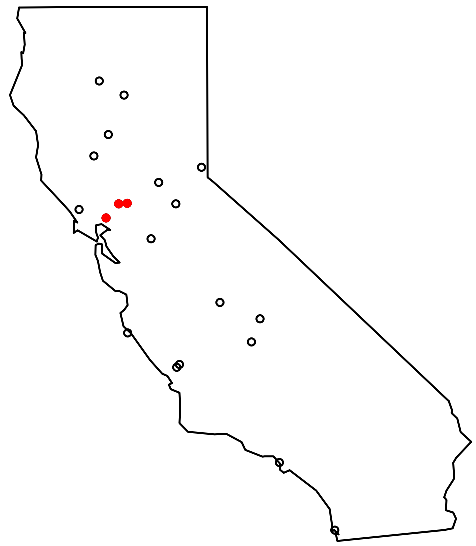

Figure 2: Locations of the 20 monitoring stations in California. The three sites in solid circles are Napa, Winters, and Davis.

37.62, 0.02). The marginal GEV model at site $s$ has parameters

$$\begin{cases} \mu(s) = \beta_{\mu,0} + \beta_{\mu,1}X_1(s) + \beta_{\mu,2}X_2(s) + \beta_{\mu,3}X_3(s), \\ \sigma(s) = \beta_{\sigma,0} + \beta_{\sigma,1}X_1(s) + \beta_{\sigma,2}X_2(s) + \beta_{\sigma,3}X_3(s), \\ \xi(s) = \beta_{\xi,0}. \end{cases}$$

This is the same model as that in Shang et al. (2011) except that the Southern Oscillation Index term is removed, which otherwise would introduce temporal nonstationarity at all sites. The spatial dependence structure is modeled by a Smith model with dispersion matrix $\Sigma$.

To estimate the parameters with the two-step approach, we chose the 95th sample percentile as the threshold $u(s)$ for each site $s$. As in standard univariate extreme value analyses, there was temporal dependence in daily precipitation and we needed to remove clustering from the observed data before applying the two-step approach. A simple way of declustering is to define consecutive exceedances of a threshold to belong to the same cluster; the cluster is terminated once an observation falls below $u(s)$ (e.g., Coles, 2001).

Table 5 summarizes the parameter estimates and their standard errors from the two-step approach and the pairwise likelihood approach. Both approaches gave qualitatively the same results, which are consistent with those in Shang et al. (2011). As the covariates were centered at the values at San Francisco, the intercepts for the location and scale parameters are interpreted as the location and scale, respectively, for the GEV distribution in San Francisco. The estimated shape parameter are positive and significantly away from zero,

Table 5: Summaries of parameter estimates and the standard errors for two approaches ($M_1$: composite pairwise likelihood approach using block maxima data; $M_2$: two-step approach using data fusion).

| Model Parameter | | $M_1$ Estimate | $M_1$ Standard Error | $M_2$ Estimate | $M_2$ Standard Error |
|---|---|---|---|---|---|
| *Marginal GEV* | | | | | |
| Location | Intercept | 4.677 | 0.167 | 5.692 | 0.073 |
| | Latitude | −0.198 | 0.049 | −0.181 | 0.026 |
| | Longitude | −0.675 | 0.056 | −0.797 | 0.032 |
| | Elevation | 0.150 | 0.017 | 0.166 | 0.017 |
| Scale | Intercept | 2.307 | 0.127 | 2.372 | 0.069 |
| | Latitude | −0.145 | 0.052 | −0.129 | 0.024 |
| | Longitude | −0.358 | 0.067 | −0.358 | 0.031 |
| | Elevation | 0.061 | 0.018 | 0.059 | 0.011 |
| Shape | Intercept | 0.173 | 0.030 | 0.092 | 0.016 |
| *Dependence structure* | | | | | |
| | $\sigma_{11}$ | 0.766 | 0.085 | 0.301 | 0.036 |
| | $\sigma_{12}$ | −1.382 | 0.153 | −0.494 | 0.061 |
| | $\sigma_{22}$ | 2.538 | 0.274 | 0.882 | 0.106 |

indicating that the tails of the marginal GEV distributions are heavier than that of a Gumbel distribution. Sites in the west and south, with higher elevation tend to have higher location parameter and higher scale parameter (higher variation). The spatial dependence among extreme precipitations is captured by the parameters in $\Sigma$. The estimated $\sigma_{11}$ is about a third of the estimated $\sigma_{22}$, indicating anisotropy in the strengths of the spatial dependence. The $\sigma_{12}$ is significantly negative, suggesting a rotation in the ellipsoidal dependence along the coast line.

The two-step approach yielded marginal parameter estimates that are of similar size to those from the pairwise likelihood approach except for the shape parameter; the shape parameter estimates are 0.173 and 0.092 from the pairwise likelihood approach and the two-step approach, respectively. The dependence parameter estimates from the two-step approach are about a third to a half of those from the pairwise likelihood approach. For this application, We observe more difference in dependence parameter estimates than in marginal parameter estimates between the two approaches. As far as the standard errors are concerned, the two-step approach gave much smaller standard errors than the pairwise likelihood approach except for one parameter — the coefficient of elevation in the location model. Most of the reductions are about a half or more.

What are the implications of the reduced standard error in inferences? Of course, reduced

Table 6: Joint 50-year return levels (cm) for three pairs based on both approaches ($M_1$: pairwise likelihood approach using block maxima data; $M_2$: two-step approach using data fusion).

| | $M_1$ | | $M_2$ | |
|---|---|---|---|---|
| Pair | 95% CI | Width | 95% CI | Width |
| Napa & Winters | (8.84, 10.82) | 1.98 | (10.63, 12.17) | 1.54 |
| Napa & Davis | (8.83, 10.78) | 1.95 | (10.62, 12.20) | 1.58 |
| Winters & Davis | (10.84, 13.61) | 2.77 | (12.34, 14.03) | 1.69 |

standard errors in marginal parameters lead to more precise inference about marginal risk measures such as return level for each site. Since our context is spatial extremes modeling, it is of interest to see how the reduced standard errors affects risk measures of jointly defined events. The first one is the joint 50-year return level for two sites $s_1$ and $s_2$, which is defined as $y_{50}$ such that $\Pr\big(Y(s_1) > y_{50}, Y(s_2) > y_{50}\big) = 1/50$. It is exceeded simultaneously at the two sites once every 50 year. Because the bivariate marginal densities of the max-stable process are known, $y_{50}$ can be found numerically for any given parameter vector. The second one is the joint sampling distribution of the maximum of the extremal precipitation over every 50 years at multiple sites. From the max-stability property, the dependence structure of the joint distribution is characterized by the same max-stable process up to a scale. Realizations from the distribution can be drawn for any number of sites and can be used to assess the joint behavior.

Consider three stations near the Sacramento area: Napa (122.25°W, 38.27°N), Winters (121.97°W, 38.52°N), and Davis (121.78°W, 38.53°N). We generated model parameters from the multivariate normal approximation of the estimator, with a large number ($N = 5000$) of realizations. For each generated parameter vector, the joint 50-year return level was obtained numerically for each pair of sites, since the bivariate marginal distribution function at two sites is known for the Smith model. Table 6 shows the 95% confidence intervals of joint 50-year return levels for three pairs with estimator from both approaches. As can be seen, the two-step approach predicts higher joint return levels than the pairwise likelihood approach but the intervals still overlap. The two-step approach always gives a tighter confidence interval, about 61–81% of the lengths from the pairwise likelihood approach. In particular, the joint return level for Winters and Davis are much higher than those for the other two pairs, which could be explained by the closeness in distance and thus strong dependence between them. The two-step approach gives an interval 39% percent narrower that the pairwise likelihood approach for this pair.

For all the three sites, we now look at the joint sampling distribution of the maximum of every 50-year period. For each of the 5000 parameter vector drawn from the asymptotic normal distribution, we generate one realization for the 50-year sitewise maxima. Figure 3 shows the scatter plot of the 5000 draws of 50-year maxima for each pair based on both

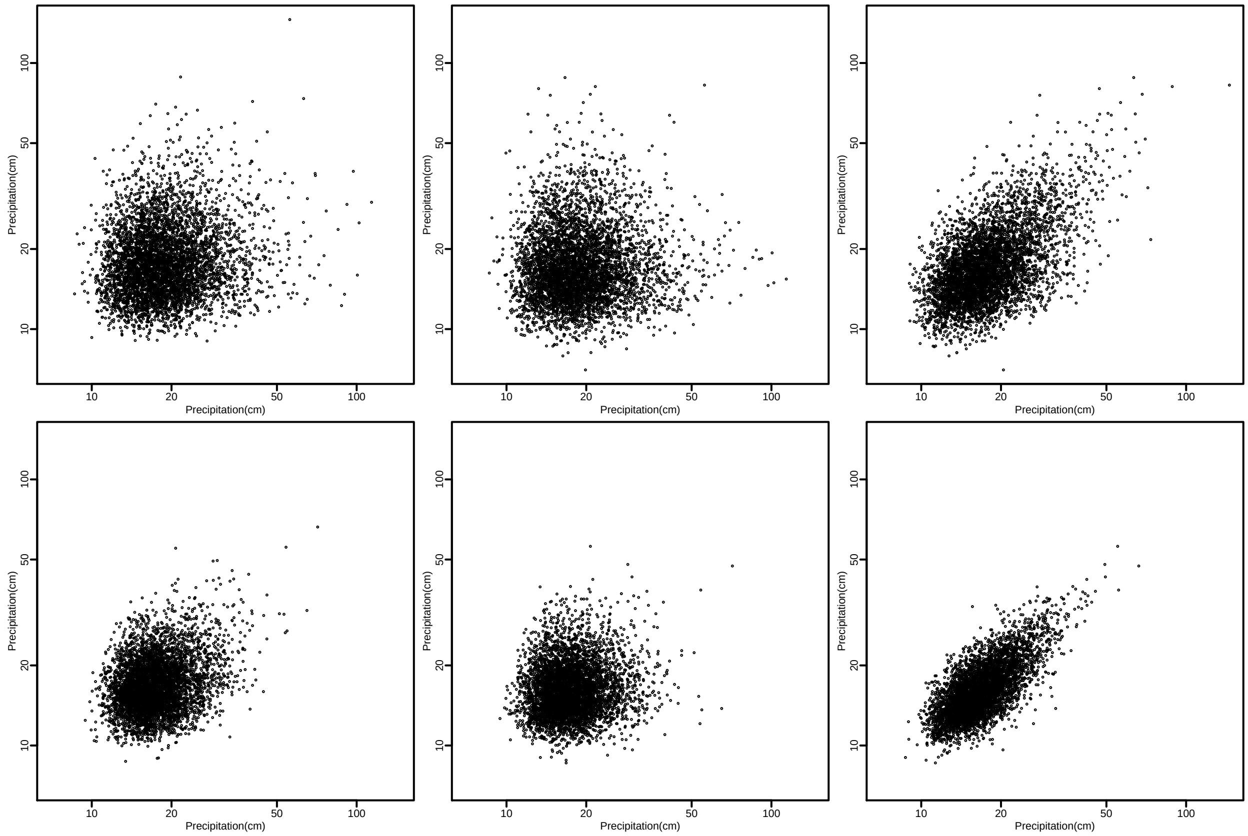


Figure 3: The 50-year sample return levels (cm) for three pairs on the log scale. Upper: pairwise likelihood approach using block maxima data; Lower: two-step approach using data fusion. Left: Napa & Winters; Center: Napa & Davis; Right: Winters & Davis.

approaches on the log scale. The two-step approach always gives a more concentrated scatter plot than the pairwise likelihood approach, thus giving a more accurate prediction. There seems to be a positive dependence between each pair, which is especially obvious for the last pair (Winters and Davis), because of their strong dependence.

# 6 Discussion

In contrast to the existing composite likelihood approach which utilizes only block maxima, our two-step approach fuses block maxima data with daily records and makes more efficient inferences about the parameters. The first step estimates marginal parameters based on an independent marginal likelihood from the point processes approach for univariate extreme value analysis; the second step estimates dependence parameters using a pairwise composite likelihood by block maximum data, with marginal parameters replaced by their estimates from the first step. This method is simple, intuitive for practitioners who are

familiar with univariate extreme value modeling, avoiding defining multivariate thresholds and multivariate Pareto process modeling (Aulbach and Falk, 2012). Our simulation study showed substantial efficiency gain from the existing composite likelihood approach to the two-step approach with realistic number of sites and sample sizes. In an application to extreme winter precipitation in California at 20 sites over 55 years, the two-step estimator gives much smaller standard errors for all parameters as well as tighter confidence intervals for joint risk measures than the pairwise likelihood approach.

Further research is merited in several aspects to extend the method. In the first step, we did not address threshold selection, an important and still active problem even for univariate extreme value analysis (e.g., Guillou and Hall, 2001; Thompson et al., 2009). Recent research showed promising approach with quantile regression for nonstationarity with covariate information (Northrop and Jonathan, 2011). Alternatively, we may use the $r$ largest order statistic to construct the marginal likelihood (e.g., Coles, 2001). This way, there is no need to specify the threshold and temporal nonstationarity may be incorporated to detect change in trend. In the second step, for max-stable process models that have known trivariate marginal densities, a triplewise likelihood can be used in place of the pairwise likelihood; for the Smith model, efficiency gain has been shown by Genton et al. (2011). More broadly, an iterative approach may be developed by adding a likelihood term involving the dependence parameters to the marginal likelihood in the first step to borrow strength from other sites using the spatial dependence in estimating the marginal parameters. How to appropriately weight the pieces in the composite likelihood and its practical utility need to be investigated.

# Acknowledgements

This research was partially supported by a contract with Environment Canada.

# A Alternative Formula for $A_n$

We first look into the contribution of each site and each pair at each year to the terms in $A_n$. Let $\psi_{1t,s}(\beta) = \partial \ell_{1t,s}/\partial\beta$. Let $\ell_{2t,(i,j)}(\theta,\beta) = \log f_{i,j}\big((M_{i,t}, M_{j,t});\theta,\beta\big)$ and $\psi_{2t,(i,j)}(\beta,\theta) = \partial\ell_{2t,(i,j)}/\partial\theta$. Then, $A_n$ can be rewritten as

$$A_n = \frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} \sum_{i=1}^{S}\partial\psi_{1t,s}(\hat\beta_n)/\partial\beta^\top & \\ \sum_{i,j=1;i<j}^{S}\partial\psi_{2t,(i,j)}(\hat\beta_n,\hat\theta_n)/\partial\beta^\top & \sum_{i,j=1;i<j}^{S}\partial\psi_{2t,(i,j)}(\hat\beta_n,\hat\theta_n)/\partial\theta^\top \end{pmatrix}.$$

Instead of calculating the second-order derivatives in $A_n$, we use the first-order derivatives based on the second Bartlett identity, assuming that the univariate and bivariate marginal models are correctly specified. Let $\phi_{2t,(i,j)}(\beta,\theta) = \partial\ell_{2t,(i,j)}/\partial\beta$. We can estimate $A$ by

$$\hat A_n = -\frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} \sum_{i=1}^{S}\psi_{1t,s}(\hat\beta_n)\psi_{1t,s}^\top(\hat\beta_n) & \\ \sum_{i,j=1;i<j}^{S}\psi_{2t,(i,j)}(\hat\beta_n,\hat\theta_n)\phi_{2t,(i,j)}^\top(\hat\beta_n,\hat\theta_n) & \sum_{i,j=1;i<j}^{S}\psi_{2t,(i,j)}(\hat\beta_n,\hat\theta_n)\psi_{2t,(i,j)}^\top(\hat\beta_n,\hat\theta_n) \end{pmatrix}.$$

# References


Aulbach, S. and Falk, M. (2012), “Testing for a Generalized Pareto Process,” *ArXiv e-prints*.

Bacro, J.-N. and Gaetan, C. (2012), “A Review on Spatial Extreme Modelling,” in *Advances and Challenges in Space-time Modelling of Natural Events*, eds. Porcu, E., Montero, J.-M., and Schlather, M., Springer, vol. 207 of *Lecture Notes in Statistics*, pp. 103–124.

Balkema, A. A. and de Haan, L. (1974), “Residual Life Time at Great Age,” *The Annals of Probability*, 2, 792–804.

Buishand, T. A., de Haan, L., and Zhou, C. (2008), “On Spatial Extremes: With Application to a Rainfall Problem,” *Annals of Applied Statistics*, 2, 624–642.

Coles, S. (1993), “Regional Modelling of Extreme Storms via Max-Stable Processes,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 55, 797–816.

— (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag Inc.

Davison, A. C. and Gholamrezaee, M. M. (2012), “Geostatistics of extremes,” *Proceedings of the Royal Society A*, forthcoming.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012), “Statistical modelling of spatial extremes (with discussion),” *Statistical Science*, forthcoming.

de Haan, L. (1984), “A Spectral Representation for Max-stable Processes,” *The Annals of Probability*, 12, 1194–1204.

de Haan, L. and Pereira, T. T. (2006), “Spatial Extremes: Models for the Stationary Case,” *The Annals of Statistics*, 34, 146–168.

Falk, M. and Guillou, A. (2008), “Peaks-over-threshold Stability of Multivariate Generalized Pareto Distributions,” *Journal of Multivariate Analysis*, 99, 715–734.

Falk, M., Hüsler, J., and Reiss, R.-D. (2010), *Laws of Small Numbers: Extremes and Rare Events*, Springer Basel.

Falk, M. and Michel, R. (2009), “Testing for a Multivariate Generalized Pareto Distribution,” *Extremes*, 12, 33–51.

Genton, M. G., Ma, Y., and Sang, H. (2011), “On the Likelihood Function of Gaussian Max-stable Processes,” *Biometrika*, 98, 481–488.

Godambe, V. P. (1991), *Estimating Functions*, Oxford University Press.

Guillou, A. and Hall, P. (2001), “A Diagnostic for Selecting the Threshold in Extreme Value Analysis,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 293–305.

Heagerty, P. J. and Lele, S. R. (1998), "A Composite Likelihood Approach to Binary Spatial Data," *Journal of the American Statistical Association*, 93, 1099–1111.

Heagerty, P. J. and Lumley, T. (2000), "Window Subsampling of Estimating Functions with Application to Regr ession Models," *Journal of the American Statistical Association*, 95, 197–211.

Joe, H. (2005), "Asymptotic Efficiency of the Two-stage Estimation Method for Copula-based Models," *Journal of Multivariate Analysis*, 94, 401–419.

Katz, R. W., Parlange, M. B., and Naveau, P. (2002), "Statistics of Extremes in Hydrology," *Advances in Water Resources*, 25, 1287–1304.

Leadbetter, M. R., Rootzen, H., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag Inc.

Lindsay, B. G. (1988), "Composite Likelihood Methods," in *Statistical Inference from Stochastic Processes*, ed. Prabhu, N. U., American Mathematical Society, pp. 221–239.

Northrop, P. J. and Jonathan, P. (2011), "Threshold Modelling of Spatially Dependent Non-Stationary Extremes with Application to Hurricane-Induced Wave Heights," *Environmetrics*, 22, 799–809.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010), "Likelihood-based Inference for Max-stable Processes," *Journal of the American Statistical Association*, 105, 263–277.

Patz, J. A., Campbell-Lendrum, D., Holloway, T., and Foley, J. A. (2005), "Impact of Regional Climate Change on Human Health," *Nature*, 438, 310–317.

Patz, J. A., Githeko, A. K., McCarty, J. P., Hussein, S., Confalonieri, U., and de Wet, N. (2003), "Climate Change and Infectious Diseases," in *Climate Change and Human Health: Risks and Responses*, eds. McMichael, A. J., Campbell-Lendrum, D. H., Corvalán, C. F., Ebi, K. L., Githeko, A. K., Scheraga, J. D., and Woodward, A., Geneva, Switzerland: World Health Organization, pp. 103–132.

Pickands, J. (1971), "The Two-dimensional Poisson Process and Extremal Processes," *Journal of Applied Probability*, 8, 745–756.

— (1975), "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, 3, 119–131.

Ribatet, M. (2011), *SpatialExtremes: Modelling Spatial Extremes*, r package version 1.8-1.

Rootzén, H. and Tajvidi, N. (2006), "Multivariate Generalized Pareto Distributions," *Bernoulli*, 12, 917–930.

Schlather, M. (2002), "Models for Stationary Max-stable Random Fields," *Extremes*, 5, 33–44.

Shang, H., Yan, J., and Zhang, X. (2011), "El Niño–Southern Oscillation Influence on Winter Maximum Daily Precipitation in California in a Spatial Model," *Water Resources Research*, 47, W11507.

Smith, R. L. (1989), "Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-level Ozone (C/R: P378-393)," *Statistical Science*, 4, 367–377.

— (1990), "Max-stable Processes and Spatial Extremes," Unpublished manuscript, University of Surrey.

Tanaka, S. and Takara, K. (2002), "A Study on Threshold Selection in POT Analysis of Extreme Floods," in *The Extremes of the Extremes: Extraordinary Floods*, eds. Snorasson, A., Finnsdottir, H., and Moss, M., IAHS Publication, vol. 271, pp. 299–304.

Thompson, P., Cai, Y., Reeve, D., and Stander, J. (2009), "Automated Threshold Selection Methods for Extreme Wave Analysis," *Coastal Engineering*, 56, 1013–1021.

Zhang, X., Wang, J., Zwiers, F. W., and Groisman, P. Y. (2010), "The Influence of Large Scale Climate Variability on Winter Maximum Daily Precipitation over North America," *Journal of Climate*, 23, 2902–2915.

Zhao, Y. and Joe, H. (2005), "Composite Likelihood Estimation in Multivariate Data Analysis," *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 33, 335–356.